INTERNATIONAL CENTRE FOR ECONOMIC RESEARCH
ICER

WORKING PAPER SERIES

Ramsés H. Mena and Luis E. Nieto-Barajas

EXCHANGEABLE CLAIM SIZES
IN A COMPOUND POISSON TYPE PROCESS

Working Paper no. 19/2007
October 2007

APPLIED MATHEMATICS AND QUANTITATIVE METHODS
WORKING PAPER SERIES

Exchangeable claim sizes in a compound Poisson type process

Ramsés H. Mena[1]* and Luis E. Nieto-Barajas[2] †

[1] *Department of Probability and Statistics, IIMAS–UNAM, 01000 Mexico, D.F.*

[2] *Department of Statistics, ITAM, Rio Hondo 1, Progreso Tizapan, 01080 Mexico, D.F.*

ABSTRACT. When dealing with risk models the typical assumption of independence among claim size distributions is not always satisfied. Here we consider the case when the claim sizes are exchangeable and study the implications when constructing aggregated claims through compound Poisson type processes. In particular, exchangeability is achieved through conditional independence and using parametric and nonparametric measures for the conditioning distribution. A full Bayesian analysis of the proposed model is carried out to illustrate.

Key words: Bayes nonparametrics, compound Poisson process, exchangeable claim process, exchangeable sequence, risk model.

1. Introduction

In insurance and risk modelling, the compound Poisson process has played a fundamental role. The compound Poisson process (CPP), denoted as $\{X_t;\, t \geq 0\}$, can be defined by

$$X_t = \sum_{i=1}^{N_t} Y_i, \tag{1}$$

where $\{N_t; t \geq 0\}$ is a homogeneous Poisson process with intensity $\lambda > 0$ and $Y_1, Y_2, \ldots$ a sequence of independent and identically distributed (nonnegative) random variables with a common density f, independent of N_t. In insurance, $\{N_t\}$ is to be interpreted as the number of claims of the company during the time interval $(0, t]$, and Y_i the magnitude of the i-th claim. Therefore X_t can be seen as the total amount of claims, or the aggregated claims, on $(0, t]$. Provided second order moments for the claim distribution exist, some basic characteristics of (1) are at order

$$\begin{aligned} \mathrm{E}(X_t) &= \lambda t\, \mathrm{E}(Y_i) \\ \mathrm{Var}(X_t) &= \lambda t\, \mathrm{E}(Y_i^2) \\ \mathrm{Cov}(X_t, X_s) &= \mathrm{Var}(X_{t\wedge s}), \end{aligned} \tag{2}$$

where $t \wedge s = \min(s, t)$. Moreover, $\mathrm{Corr}(X_t, X_s) = (t \wedge s)/\sqrt{ts}$, so if $t < s$ then $\mathrm{Corr}(X_t, X_s) = \sqrt{t/s}$.

*Ramsés H. Mena is grateful for the support from the International Centre for Economic Research (ICER), Turin, Italy.

†Supported by grant J48072-F from CONACYT, Mexico.

In the collective risk theory it is common to model the risk of a company $\{Z_t;\, t \geq 0\}$ through the CPP in the following way

$$Z_t = r\, t - X_t, \tag{3}$$

where $r > 0$ is the constant *gross risk premium rate*. This model constitutes the classical risk model introduced by Lundberg (1926) and further analyzed by Cramér (1930). In the case of an insurance company, r denotes the units of money received by the company per unit of time. Hence, Z_t is the profit of the company over the time interval $(0, t]$.

One of the main interest regarding the wealth of a company, with initial capital u and facing risk process (3), is concerning the *ruin probability* $\Psi(u)$ defined through

$$\Psi(u) := \mathrm{P}\left(\inf_{t\geq 0}\{t; u + Z_t < 0\} < \infty\right) \tag{4}$$

or through its *finite-horizon* analog

$$\Psi(u, T) := \mathrm{P}\left(\inf_{0\leq t<T}\{t; u + Z_t < 0\} < T\right) \tag{5}$$

Such quantities are relevant when, in average, a company is facing a profitable business. This feature is measured through the *relative safety loading* ρ, which represents the expected profit/loss relative to the total amount of claims, that is

$$\rho := \frac{\mathrm{E}(Z_t)}{\mathrm{E}(X_t)}.$$

Hence an insurance company is willing to have a small ruin probability $\Psi(u)$ when starting a business with $\rho > 0$.

There are several generalizations to the simple compound Poisson process (1), most of them inherited by a generalization of the underlying Poisson process $\{N_t\}$. For example, $\{N_t\}$ could be replaced by a *mixed Poisson process* which is defined as a Poisson process with random intensity, namely $N_t \mid \Delta \sim \mathrm{Po}(t\Delta)$ and $\Delta \sim G$ for some distribution G supported on $\mathbb{R}^+$. An even more general approach could be undertaken by considering $\{N_t\}$ a *Cox process* which instead of considering Δ invariant over time, considers a random measure $\Delta := \{\Delta(t); t \geq 0\}$ with realizations belonging to the set of non-decreasing right-continuous finite measures $\mathcal{M}$. Namely, $N_t \mid \Delta \sim \mathrm{Po}(\Delta(t))$ and $\Delta(t)$ being a realization of a random measure $\mathscr{P}$. Clearly these approaches might result in processes with increments no longer being independent. For instance, it can be seen that a stationary mixed Poisson process $\{N_t\}$ (and then the resulting compound process X_t) have exchangeable increments (Daboni, 1974). For mixed Poisson and Cox processes see Grandell (1997). Other generalizations of the risk process (3) include the work of Morales and Schoutens (2003), who considered a Lévy process for modeling the aggregate claims X_t.

The main objective of this paper is to generalize the compound Poisson processes (1) by relaxing the independence assumption in the claims, but keeping the counting process N_t to be a homogeneous Poisson process. The idea of imposing dependence among claims has been considered before by other authors, for instance: Gerber (1982) and Promislow (1991) who imposed a dependence structure through a linear ARMA model; Cossette and Marceau (2000) modeled dependence through a Poisson shock model and studied its implications in the ruin probability, and Mikosch and Samorodnitsky (2000) who model the claims through a stationary stable process and also investigated the impact on ruin probabilities.

The approach we undertake in this paper is based on a dependence structure provided via exchangeability assumptions in the claims. It is worth noting that exchangeability has also been used to generalize risk models but in different ways. Bühlmann (1960, 1970) and Daboni (1974) generalized the counting process and used a mixed Poisson process specially designed to have exchangeable interarrival times. In addition, we use a particular construction that allows us to have dependence among claims while keeping the marginal distribution invariant. The outline of the paper is as follows. In Section 2 we define the exchangeable claims process and provide two ways of defining exchangeable sequences through a parametric and a nonparametric approaches. Section 3 describes a methodology to implement Bayesian inference for the parameters in both the traditional compound Poisson process and the proposed approach. Finally in Section 4 we illustrate our model with a full Bayesian analysis of catastrophic claims.

2. Exchangeable claims modeling

Definition 1. Let $\{N_t; t \geq 0\}$ be a simple Poisson process with intensity $\lambda > 0$ and $Y_1, Y_2, \ldots$ a sequence of exchangeable nonnegative random variables with common one-dimensional marginal distribution F, independent of $\{N_t\}$. Then the *exchangeable claim process* (ECP), $\{X_t; t \geq 0\}$, is defined as

$$X_t := \sum_{i=1}^{N_t} Y_i \tag{6}$$

Clearly this new process has no longer independent increments, however it is still a Markov process with exchangeable increments. Specific cases of similar summations of the type defined by (6) have been previously used to generalize the central limit theorem for exchangeable sequences, see for example Klass and Teicher (1987).

When analysing real data it is of interest to have an arbitrary but given claim distribution, therefore the issue here is how to construct an exchangeable sequence with given marginal distributions. Clearly the easiest case is that of iid claims, however the point here is to allow the model to have a possible dependence structure among the claims.

Based on the representation theorem of de Finetti (1937), we can construct a sequence of exchangeable random variables through a conditional independence sequence. More explicitly, the random variables $Y_1, Y_2, \ldots$ are an exchangeable sequence if there exists a parameter/measure G such that $\{Y_i\}$ for $i = 1, 2, \ldots$ are conditionally independent given G, and G is a random parameter/measure (known as de de Finetti's measure) with law described by a distribution/process $\mathcal{P}$. Note that exchangeable variables defined in this fashion are always positive correlated, which suffices for most dependencies found in risk modeling.

Proposition 1. *Let $\{X_t;\, t \geq 0\}$ be a ECP with $Y_i \mid G \sim G$, $i = 1, 2, \ldots$ conditionally independent given G, and $G \sim \mathcal{P}$, then assuming the existence of second order marginal moments we have:*

(i) $E(X_t) = \lambda t E(Y_i)$

(ii) $Var(X_t) = \lambda t E(Y_i^2) + \lambda^2 t^2 Cov(Y_i, Y_j)$

(iii) $Cov(X_t, X_s) = \lambda\,(t \wedge s)\, E[Y_i^2] + \lambda^2\, ts\; Cov(Y_i, Y_j)$

Proof. The whole idea is based on the conditional independence properties.

(i)

$$\mathrm{E}(X_t) = \mathrm{E}\left\{\mathrm{E}\left(\sum_{i=1}^{N_t} Y_i \,|N_t\right)\right\} = \mathrm{E}\,(N_t)\;\mathrm{E}\,(Y_i) = \lambda\, t\, \mathrm{E}(Y_i),$$

where the second equality follows due to the independence between N_t and Y_i.

(ii) First, let us notice that

$$\mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}\left\{\mathrm{E}(Y_i \mid G), \mathrm{E}(Y_j \mid G)\right\} = \mathrm{Var}\{\mathrm{E}(Y_i \mid G)\},$$

hence

$$\begin{aligned}
\mathrm{Var}(X_t) &= \mathrm{E}\{\mathrm{Var}(X_t \mid G)\} + \mathrm{Var}\{\mathrm{E}(X_t \mid G)\} \\
&= \lambda\, t\, \mathrm{E}(Y_i^2) + \lambda^2\, t^2\, \mathrm{Var}\{\mathrm{E}(Y_i \mid G)\} \\
&= \lambda\, t\, \mathrm{E}(Y_i^2) + \lambda^2\, t^2\, \mathrm{Cov}(Y_i, Y_j)
\end{aligned}$$

(iii)

$$\begin{aligned}
\mathrm{Cov}(X_t, X_s) &= \mathrm{Cov}\left\{\mathrm{E}\left(X_t \mid G\right), \mathrm{E}\left(X_s \mid G\right)\right\} + \mathrm{E}\left[\mathrm{Cov}(X_t, X_s \mid G)\right] \\
&= \mathrm{Cov}\left\{\lambda t\, \mathrm{E}(Y_i \mid G), \lambda s\, \mathrm{E}(Y_j \mid G)\right\} + \mathrm{E}\{\mathrm{Var}(X_{t\wedge s} \mid G)\} \\
&= \lambda^2\, ts\, \mathrm{Cov}(Y_i, Y_j) + \lambda\, (t \wedge s)\, \mathrm{E}[\mathrm{Y}_i^2].
\end{aligned}$$

If we define processes CPP and ECP with the same one-dimensional marginal claims distribution, then comparing their characteristics we note that both processes have the same expected value. However, the variance and covariance are not the same; the variance (covariance) of the ECP is larger than the variance (covariance) of the CPP.

In terms of the correlation, if we let $h(t) := \mathrm{Cov}(Y_i, Y_j)\lambda t + \mathrm{E}[Y_i^2]$ and $t < s$ then $\mathrm{Corr}(X_t, X_s)$ $= \sqrt{\frac{t}{s}} \frac{h(s)}{\sqrt{h(s)h(t)}}$, which means that the correlation in the ECP is also larger. This behavior is later illustrated in Examples 1 and 2.

It is worth mentioning that in some applications, where heavy tailed distributions are used for claim modeling, the covariance or correlation might not be a good indicator of dependence since typically when using such distributions second moments do not exist, hence one has to resort to other measures of dependence such as Kendall's Tau or Spearman's Rho.

Regarding the classical risk model (3) and thanks to a well-known renewal argument (Feller, 1971), the non-ruin probability $\varphi(u) := 1 - \Psi(u)$ for the CPP process satisfies $\varphi(u) = E[\varphi(u - rT_1 - X_1)]$, where $T_1 \sim \mathrm{Exp}(\lambda)$ denotes the time of the first claim and $X_1 \sim F$. Hence, it can be proved that

$$\varphi(u) = \frac{\lambda}{r} e^{\lambda u/r} \int_u^\infty e^{-\lambda x/r} \int_0^x \varphi(x - z) F(dz) dx \tag{7}$$

Noting that conditionally on G, the ECP is a CCP, and since $F \mapsto \varphi_F$ is a linear functional, we have that for the ECP

$$\mathrm{E}[\varphi \mid G] = \varphi_F,$$

provided $\mathrm{E}[G] = F$. Therefore, at the marginal level, the ruin probability for the ECP coincides with that of the CPP. However the effects on the finite-horizon ruin probability of the overdispersion found in the moment-properties underlying to the ECP, are that this latter process ruins faster than the CPP, namely for a fixed T, $\Psi_{ECP}(u, T) \geq \Psi_{CCP}(u, T)$.

Now we will consider two ways of defining exchangeable sequences of random variables with given marginal distributions, through a parametric and a nonparametric conditioning distributions.

2.1. Exchangeable sequences: Parametric method

The idea is to define an exchangeable sequence $Y_1, Y_2, \ldots$ in such a way that each Y_i will have the same marginal distribution $F(y)$. For the sake of exposition we assume the existence of a density $f(y)$. For that, we introduce a latent variable Z with arbitrary conditional density $f(z \mid y)$. Then, we define $f(y \mid z)$ using Bayes' Theorem in the following way:

$$f(y \mid z) = \frac{f(z \mid y) f(y)}{f(z)} \tag{8}$$

with

$$f(z) = \int f(z \mid y) f(y) \mathrm{d}\mu_1(y), \tag{9}$$

where $\mu_1(y)$ represents a reference measure such as counting measure if Y is discrete or the Lebesgue measure if Y is continuous. It is straightforward to show that if we marginalize over Z then,

$$\int f(y \mid z) f(z) \mathrm{d}\mu_2(z) = f(y)$$

as required, where $\mu_2(\cdot)$ is another reference measure acting on Z. Then, if we take $Y_i \mid Z \sim f(y \mid z)$, as in (8), for $i = 1, 2, \ldots$ a sequence of conditional independent random variables given $Z = z$, with marginal distribution for Z, as in (9), then $Y_1, Y_2, \ldots$ is a sequence of exchangeable random variables with marginal densities $f(y)$. A similar idea was used by Pitt *et al.* (2002) to construct stationary autoregressive models with given marginal distributions, although in their construction different Z_i's are used for different Y_i's.

Notice that the possibilities for Z are extensive, for example it could be discrete, continuous, univariate or multivariate. Hence different features of Z will lead to different forms of dependence.

Example 1. Let us denote by $\mathrm{Ga}(a, b)$ a gamma distribution with mean a/b, by $\mathrm{Po}(\lambda)$ a Poisson distribution with mean λ and by $\mathrm{Pga}(a, b, c)$ a Poisson–gamma distribution with mean $(c\,a)/b$. We will define an exchangeable sequence with $\mathrm{Ga}(a, b)$ as marginal distribution by assuming $f(z \mid y) = \mathrm{Po}(z \mid c\,y)$, $c > 0$. In this case we obtain that, $f(y \mid z) = \mathrm{Ga}(y \mid a + z, b + c)$ and $f(z) = \mathrm{Pga}(z \mid a, b, c)$. Hence, if we take $Y_i \mid Z \sim \mathrm{Ga}(a + z, b + c)$ for $i = 1, 2, \ldots$ conditionally independent given Z and $Z \sim \mathrm{Pga}(a, b, c)$ hence marginally $Y_i \sim \mathrm{Ga}(a, b)$ with $\mathrm{Corr}(Y_i, Y_j) = c/\{(b + c)\}$ for all $i \neq j$. Now, since $N_t \sim \mathrm{Po}(\lambda t)$ is the number of claims in $(0, t]$ and $Y_1, Y_2, \ldots$ the claim sizes such that $Y_i \sim \mathrm{Ga}(a, b)$, then $X_t = \sum_{i=1}^{N_t} Y_i$ is the total amount of claims in $(0, t]$. We will consider two processes, namely X_t^I and X_t^E, the one with independent claims and the one with exchangeable claims respectively. Using the moments in (2) and Proposition 1, it is not difficult to show that the expected value of the total amount of claims are,

$$\mathrm{E}\left\{X_t^I\right\} = \mathrm{E}\left\{X_t^E\right\} = \frac{a}{b} \lambda t.$$

The variance of the processes become,

$$\mathrm{Var}\left\{X_t^I\right\} = \frac{a(a+1)}{b^2} \lambda t$$

and

$$\mathrm{Var}\left\{X_t^E\right\} = \mathrm{Var}\left\{X_t^I\right\} + \frac{ac}{b^2(b+c)} \lambda^2 t^2,$$

which clearly shows that X_t^E is overdispersed with respect to X_t^I. Finally, the covariance function for $t < s$ takes the form,

$$\mathrm{Cov}\left\{X_t^I, X_s^I\right\} = \mathrm{Var}\left\{X_t^I\right\}$$

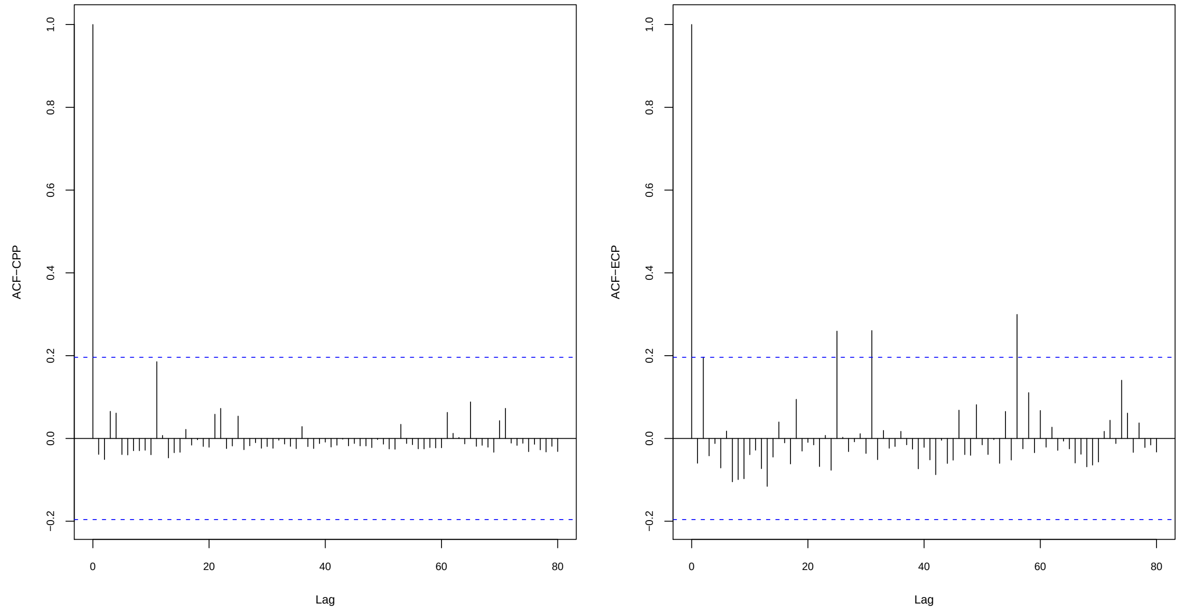

Figure 1: Autocorrelated functions (ACF) for the simulated increments $X_t - X_{t-1}$, for $t = 1, \ldots, 100$. CPP (left panel), ECP (right panel).

and

$$\text{Cov}\left\{X_t^E, X_s^E\right\} = \text{Cov}\left\{X_t^I, X_s^I\right\} + \frac{ac}{b^2(b+c)}\lambda^2 ts.$$

In order to have an idea what the implications of having an exchangeable sequence are, Figure 1 shows the autocorrelation functions corresponding to the increments $X_t - X_{t-1}$, $t = 1, \ldots, 100$, for the CPP (left panel) and the ECP (right panel) when taking $(a, b) = (0.1, 0.1)$, $c = 1$ and $\lambda = 1$. Clearly the autocorrelations for the ECP, as measures of second order dependence, are larger in average than those of the CPP process.

For both processes, the relative safety loading is the same and is given by $\rho = (r\, b)/(\lambda\, a) - 1$, which in our case, for $r = 2$, turns out to be one. However, the ECP ruins faster than the CPP process. This behavior can be seen in Figure 2 (left panel) where a realization of the surplus process, $u + Z_t$, is included for an initial capital $u = 5$. To be more precise, in Figure 2 (right panel), we also computed the crude Monte Carlo estimates of the finite-horizon ruin probabilities with $T = 100$, for initial capitals $u = 0.5, 1, 1.5, \ldots, 8$, each based on 5,000 realizations. As we can see, the ruin probability corresponding to the ECP is uniformly larger than that of the CPP, which is consistent with other studies (for example, Cossette and Marceau (2000)).

Notice that a different choice of the conditional distribution $f(z \mid y)$, in the above construction, leads to a different dependence structure. For instance, choosing $f(z \mid y) = \text{Ga}(z \mid c, y)$ implies that $f(y \mid z) = \text{Ga}(y \mid a + c, b + z)$ and $f(z) = \text{Gga}(z \mid a, b, c)$, where $\text{Gga}(z \mid a, b, c)$ denotes a gamma-gamma distribution with mean $c\, b/(a - 1)$. Therefore, for constructing an exchangeable sequence we take $Y_i \mid Z \sim \text{Ga}(a + c, b + z)$ for $i = 1, 2, \ldots$ conditionally independent given Z and $Z \sim \text{Gga}(a, b, c)$ hence $Y_i \sim \text{Ga}(a, b)$ with $\text{Corr}(Y_i, Y_j) = c/\{(a + c + 1)\}$, for $i \neq j$. $\circ$

2.2. Exchangeable sequences: Nonparametric method

The parametric method described in the previous section has the feature of having a wide variety of choices for the conditional distribution $f(z \mid y)$, leading then to different parametric

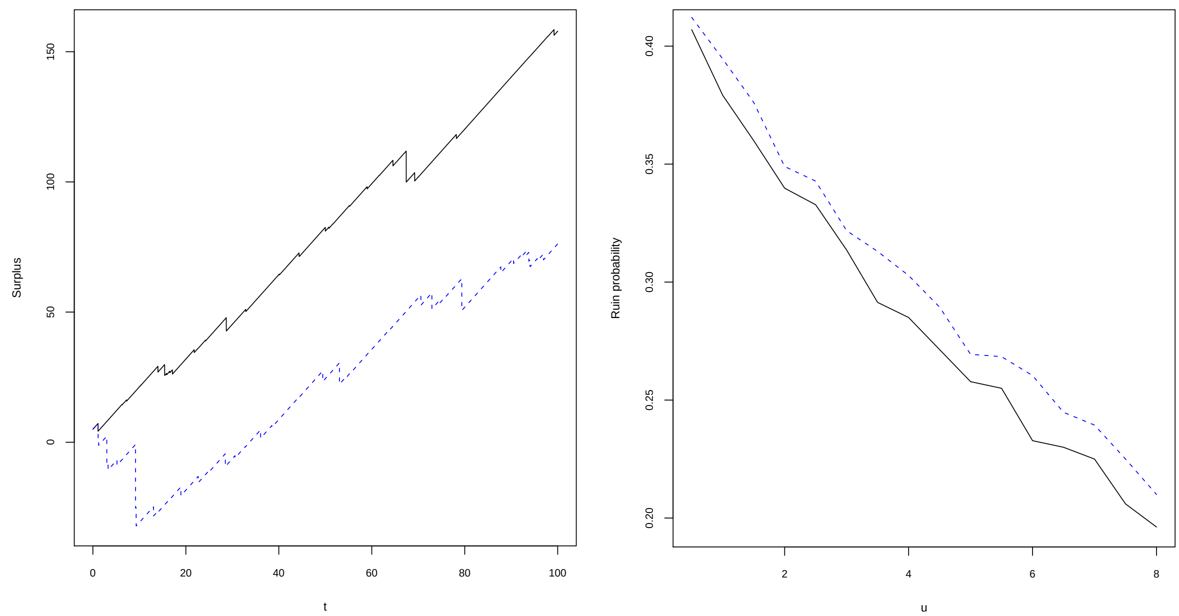

Figure 2: Surplus process (left panel) and finite-horizon ruin probabilities (right panel) for the simulated processes. CPP (solid line) and ECP (dashed line).

dependence structures. Alternatively, instead of having a latent random variable Z, we could consider a latent random distribution G, i.e., the conditional density $f(z \mid y)$ will be replaced by a nonparametric density measure $\mathcal{P}$. By doing this, we will be able to define a nonparametric dependence structure within the exchangeable sequence. The nonparametric nature underlying to this construction coveys to a dependence between the variables not depending on the marginal distributions.

Having in mind the previous parametric construction, instead of the latent variable Z we consider a latent random distribution G with conditional random distribution $G \mid Y \sim \mathcal{P}(\cdot \mid y)$. In this case,

$$F(\mathrm{d}y \mid G) = G(\mathrm{d}y)$$

with

$$G \sim \mathcal{P}.$$

Proceeding analogously to the previous construction we would need to choose $\mathcal{P}$ such that

$$\mathrm{E}_{\mathcal{P}}\{G(\mathrm{d}y)\} = \int G(\mathrm{d}y)\mathcal{P}(\mathrm{d}G) = F(\mathrm{d}y).$$

The above is a characteristic property that many probability measures on the space of distributions, used in the Bayesian nonparametric literature, satisfy. This is the case of the seminal Dirichlet process introduced by Ferguson (1973) and most of its generalizations such as species sampling models presented by Pitman (1996) and normalized random measures with independent increments analyzed in Regazzini *et al.* (2003). This characteristic, is attractive in the Bayesian nonparametric literature, since it allows to set F as the prior guess (mean) at the shape for the realizations of G.

Now, if we take $Y_i \mid G \sim G$, for $i = 1, 2, \ldots$ conditional independent random variables given G, and $G \sim \mathcal{P}$, then $Y_1, Y_2, \ldots$ is a sequence of exchangeable random variables with marginal distributions $F(\mathrm{d}y) = \mathrm{E}_{\mathcal{P}}\{G(\mathrm{d}y)\}$. Mena and Walker (2005) also used similar ideas based on nonparametric predictive distributions to define the dynamics of a first order autoregressive processes.

Example 2. In order to define an exchangeable sequence with marginal distributions $\mathrm{Ga}(a,b)$ through the nonparametric method just described, we consider a Dirichlet process $\mathcal{DP}(F/c)$ as the law $\mathcal{P}$ of G, where $c>0$ and F is a parametric c.d.f. that coincides with the mean of the process G, for details see Ferguson (1973). Now we want that $\mathrm{E}_{\mathcal{DP}}\{G(\mathrm{d}y)\} = F(\mathrm{d}y)$, with F the c.d.f. of a gamma distribution. Thus, if we take $Y_i \mid G \sim G$ for $i=1,2,\ldots$ conditionally independent given G, and $G \sim \mathcal{DP}(F/c)$, with $F(\mathrm{d}y) = \mathrm{Ga}(y \mid a,b)\mathrm{d}y$ and $c>0$ then $Y_i \sim \mathrm{Ga}(a,b)$ with $\mathrm{Cov}(Y_i,Y_j) = \mathrm{Var}_{\mathcal{DP}}(\mu_G)$ and

$$\mu_G = \int y G(\mathrm{d}y),$$

for $i \neq j$. $\circ$

The distribution of the mean μ_G has been studied by Cifarelli and Regazzini (1990) and Regazzini et al. (2002) among others. These authors provided an expression for the distribution of μ_G for any centering function F, however, moments of this distribution are not available in closed from, except for particular cases of F. On the other hand, there is an alternative way of obtaining dependence properties of an exchangeable sequence modeled by the Dirichlet process.

According to Blackwell and MacQueen (1973), the joint distribution of the exchangeable sequence $\{Y_i\}$ where $Y_i \mid G \sim G$ for $i=1,2,\ldots$ conditionally independent given G, and $G \sim \mathcal{DP}(F/c)$, can be characterized, after a marginalization of G, by

$$Y_1 \sim F,$$

$$Y_2 \mid Y_1 \sim \frac{1}{c+1}F + \frac{c}{c+1}F_1$$

and in general,

$$Y_i \mid Y_1,\ldots,Y_{i-1} \sim \frac{1}{1+c(i-1)}F + \frac{c(i-1)}{1+c(i-1)}F_{i-1}(y_i), \tag{10}$$

where $F_i(\cdot)$ is the empirical distribution function (e.d.f.) of the first $i-1$ observations. With this characterization of the exchangeable sequence, induced by the Dirichlet process, we are able to compute the covariance of any pair (Y_i,Y_j).

Proposition 2. *Let $\{Y_i\}$ be an exchangeable sequence such that $Y_i \mid G \sim G$ for $i=1,2,\ldots$ conditionally independent given G, and $G \sim \mathcal{DP}(F/c)$, where F is a centering function and $c>0$. Then,*

$$\mathit{Corr}(Y_i,Y_j) = \frac{c}{c+1}$$

for all $i \neq j$.

Proof. Let $\mu = \mathrm{E}(Y_i)$ and $\sigma^2 = \mathrm{Var}(Y_i)$, which correspond to the expected value and variance of F. Then, using conditional expectation we express $\mathrm{E}(Y_1\,Y_2) = \mathrm{E}\{Y_1 \mathrm{E}(Y_2 \mid Y_1)\}$. Now, based on the Pólya urn representation (10) we obtain that $\mathrm{E}(Y_2 \mid Y_1) = \mu/(c{+}1) + cY_1/(c{+}1)$ which implies that $\mathrm{E}(Y_1\,Y_2) = \mu^2 + \sigma^2 c/(c+1)$. Thus, $\mathrm{Cov}(Y_1,Y_2) = \sigma^2 c/(c+1)$ and $\mathrm{Corr}(Y_1,Y_2) = c/(c+1)$. Therefore, as $Y_1,Y_2,\ldots$ is an exchangeable sequence then $\mathrm{Corr}(Y_i,Y_j) = \mathrm{Corr}(Y_1,Y_2)$, which completes the proof. $\square$

Proposition 2 shows that the correlation induced in the exchangeable sequence $\{Y_i\}$ by means of the Dirichlet process is, in fact, nonparametric in the sense that it is independent of the marginal distribution of the Y_i's and only depends on the parameter c. This is in contrast with the parametric constructions of exchangeable sequences where the correlation depends on the marginal distribution of the Y_i's.

Example 2 (Continued...) We consider the two processes X_t^I and X_t^E based on independent and exchangeable claims respectively, but now exchangeability is defined through the nonparametric (Dirichlet) construction with $\text{Ga}(a,b)$ marginals. Then from Proposition 1, the moments of the total amount of claims in both cases are,

$$\mathrm{E}\left\{X_t^I\right\} = \mathrm{E}\left\{X_t^E\right\} = \frac{a}{b}\lambda t,$$

$$\mathrm{Var}\left\{X_t^E\right\} = \mathrm{Var}\left\{X_t^I\right\} + \frac{ac}{b^2(c+1)}\lambda^2 t^2,$$

and the covariance for $t < s$ is

$$\mathrm{Cov}\left\{X_t^E, X_s^E\right\} = \mathrm{Cov}\left\{X_t^I, X_s^I\right\} + \frac{ac}{b^2(c+1)}\lambda^2 ts,$$

where $\mathrm{Var}\left\{X_t^I\right\}$ and $\mathrm{Cov}\left\{X_t^I, X_s^I\right\}$ are given in Example 1. Again, overdispersion of the process X_t^E with respect to process X_t^I is also clear.

With exchangeable sequences constructed via nonparametric distributions it is possible to define a ECP with gamma marginal distributions, in particular, for the claims as in Example 2. In terms of second order moments of the process X_t^E there is not much gain in going from a parametric to a nonparametric construction in the exchangeable sequence. However, for inference purposes there is an advantage that will be illustrated in the following section, as well as gaining more flexibility in the whole distribution of the model X_t^E, which is appealing when modeling extremes events.

3. Bayesian inference of CPP and ECP

Once we have proposed a way of constructing exchangeable sequences for defining a ECP, then for a given data set we would like to make inference on the parameters of both, the CPP and the ECP models. The Bayesian approach for making inference has become very popular in several areas, including actuarial sciences (see, for example, Klugman (2003)), due to its advantage of combining all available information in a probability distribution. Here we will follow this approach.

3.1. CPP

In order to set ideas, let us start with the traditional CPP with independent claim sizes. Let X_t be a CPP as defined in (1) with $N_t \mid \lambda \sim \text{Po}(\lambda t)$ and $Y_i \sim f(y \mid \theta)$ independent for $i = 1, 2, \ldots$ and independent of N_t. Available information usually represents a realization of the process, that is, $N_t = n$ is the number of claims registered for an insurance company up to time t, and $Y_1, Y_2, \ldots, Y_n$ are the claim sizes (amounts) for the n claims.

Due to the independence of the Poisson process N_t and the claim sizes Y_i, inference for θ and λ can be done separately. Now, given $N_t = n$, the likelihood for θ is given by

$$f(y_1, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta).$$

If we assume that our prior knowledge on θ is summarized in a prior distribution $\pi(\theta)$, then the posterior distribution of θ is obtained through the Bayes' Theorem, i.e.,

$$\pi(\theta \mid \mathbf{y}) \propto f(\mathbf{y} \mid \theta)\pi(\theta).$$

Now, if we let $J_1, J_2, \ldots$ be the jump times in a $\text{Po}(\lambda t)$ process, it is well known that the inter-arrival times $W_1 = J_1, W_2 = J_2 - J_1, \ldots$ are independent variables with $\text{Ga}(1, \lambda)$ distributions. Therefore, inference on λ reduces to the standard estimation problem of the scale parameter in a sample of independent gamma random variables. Thus, a conjugate analysis for λ is obtained when assuming a prior distribution $\pi(\lambda) = \text{Ga}(\lambda \mid \alpha_\lambda, \beta_\lambda)$ and in this case the posterior distribution for λ has the form

$$\pi(\lambda \mid \mathbf{w}) = \text{Ga}\left(\lambda \,\middle|\, \alpha_\lambda + n,\, \beta_\lambda + \sum_{i=1}^{n} w_i\right), \tag{11}$$

which in terms of the J_i's can be expressed as $\pi(\lambda \mid \mathbf{J}) = \text{Ga}(\lambda \mid \alpha_\lambda + n, \beta_\lambda + J_n)$.

3.2. ECP

Lets now assume that X_t is a ECP as the one given in Definition 1, with $N_t \mid \lambda \sim \text{Po}(\lambda t)$ and $Y_1, Y_2, \ldots$ an exchangeable sequence with $f(y \mid \theta)$ marginal distributions, and the sequence of Y_i's independent of the process N_t. Let us consider two cases, the parametric and the nonparametric constructions:

Parametric case

Regarding the parametric construction of the exchangeable sequence with given marginal distributions as in Section 2.1, that is, via a latent variable $Z \sim f(z \mid c)$ and $Y_i \mid Z \overset{\text{iid}}{\sim} f(y \mid z)$. We first note that this construction implies that all Y_i's are generated given a common $Z = z$, therefore, the whole sample of y_i's provides information about a single Z and this is equivalent of having a sample of size one for Z, which is not enough to identify the value of c in the exchangeable construction. In summary, if we want to make inference on c we require at least two exchangeable sequences generated with the same mechanism.

Consider m periods of time, say $j = 1, \ldots, m$. For each period j we observe a realization of a ECP, with independence between periods. Now, given $N_{t_j} = n_j$, the number of claims up to time t_j, $\mathbf{Y}_j = (Y_{1j}, \ldots, Y_{n_j,j})$ denote the n_j exchangeable claim sizes for $j = 1, \ldots, m$. Then the likelihood for (θ, c) is given by

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_m \mid \theta, c) = \prod_{j=1}^{m} f(\mathbf{y}_j \mid \theta, c) \tag{12}$$

where

$$f(\mathbf{y}_j \mid \theta, c) = f(y_{1j}, \ldots, y_{n_j,j} \mid \theta, c) = \int \left\{ \prod_{i=1}^{n_j} f(y_{ij} \mid z_j, \theta) \right\} f(z_j \mid c) \mathrm{d}\mu_2(z_j). \tag{13}$$

The marginalization over z_j does not usually have an analytic expression, except for particular cases. When expression (13) is available analytically, inference on (θ, c) is conducted as in the CPP case via the Bayes' Theorem using the likelihood (12). However, if the integral in expression (13) is not analytically available in closed form, we can get around by considering, for the moment, that we have observed $Z_j = z_j$ along with the y_{ij}'s. If we denote $\mathbf{z} = (z_1, \ldots, z_m)$, the extended likelihood has the form

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_m, \mathbf{z} \mid \theta, c) = \prod_{j=1}^{m} f(\mathbf{y}_j, z_j \mid \theta, c)$$

with

$$f(\mathbf{y}_j, z_j \mid \theta, c) = \left\{ \prod_{i=1}^{n_j} f(y_{ij} \mid z_j, \theta) \right\} f(z_j \mid c). \tag{14}$$

If $\pi(\theta, c)$ represents our prior knowledge on (θ, c) then the posterior distribution is given by

$$\pi(\theta, c \mid \mathbf{y}, \mathbf{z}) \propto f(\mathbf{y}_1, \ldots, \mathbf{y}_m, \mathbf{z} \mid \theta, c)\pi(\theta, c).$$

Now, remembering that we have assumed that $Z_j = z_j$ was observed, to make posterior inference is easier if we implement a Gibbs sampling scheme (see, for example, Smith and Roberts, 1993) with the previous conditional posterior distribution of θ and c as well as the conditional distribution of each Z_j which is given by

$$f(z_j \mid \mathbf{y}, \theta, c) \propto f(\mathbf{y}_j, z_j \mid \theta, c)$$

given in (14), for $j = 1, \ldots, m$. By doing this, we will have a simulated value of Z_j in each iteration. Finally, due to the independence between N_{t_j} and the Y_{ij}'s, the posterior distribution for λ becomes

$$\pi(\lambda \mid \mathbf{J}) = \mathrm{Ga}\left(\lambda \,\middle|\, \alpha_\lambda + \sum_{j=1}^{m} n_j,\, \beta_\lambda + \sum_{j=1}^{m} J_{n_j} \right), \tag{15}$$

where J_{n_j} denotes the time where the last claim of period j occurred.

Nonparametric case

Let us now consider the nonparametric construction of the exchangeable sequence with given marginal distributions via the Dirichlet process as in Section 2.2. Due to the choice of the Dirichlet process to construct the exchangeable sequence, inference on θ and c will not require several realizations of the ECP as in the parametric case, only one realization of the process would be enough as long as the observed claims sizes had repeated values. Inference on c with only one exchangeable sequence is a consequence of the discreteness of the Dirichlet process (see, for example, Blackwell and MacQueen, 1973) which will be explained as follows.

Given that we have observed $N_t = n$ claims with claim sizes $Y_1, \ldots, Y_n$ then the likelihood function for (θ, c) is given by

$$f(y_1, \ldots, y_n \mid \theta, c) = \mathrm{E}_{\mathcal{DP}} \left\{ \prod_{i=1}^{n} G(\mathrm{d}y_i) \right\}.$$

Blackwell and MacQueen (1973) showed that this joint distribution can be obtained by the product of expressions coming form the Pólya urn representation (10). Therefore,

$$f(y_1, \ldots, y_n \mid \theta, c) = \prod_{i=1}^{n} \left\{ \left(\frac{1}{1 + c(i-1)} \right) f(y_i \mid \theta) + \left(\frac{c(i-1)}{1 + c(i-1)} \right) \sum_{j=1}^{i-1} \delta_{Y_j}(y_i) \right\},$$

where $\delta_Y(\cdot)$ is the degenerated measure that assigns probability one to the point Y. If we let $y_1^*, \ldots, y_k^*$ the distinct y_i's, with $k \leq n$ and after some algebra, the previous expression can be simplified to

$$f(y_1, \ldots, y_n \mid \theta, c) = \frac{(1/c)^k \Gamma(1/c)}{\Gamma(1/c + n)} \prod_{i=1}^{k} f(y_i^* \mid \theta), \tag{16}$$

where $\Gamma(\cdot)$ denotes the gamma function.

Having a closer look, we can see that as a function of θ, this likelihood is the same as the likelihood obtained with the traditional CPP but when considering the distinct observations only. This is a feature of the Dirichlet process that allows ties in the observations and thus information on θ only comes from the distinct observations. Moreover, the number of distinct observations k provides information about the parameter c, a smaller value of k produces a sharper likelihood for c; however, if no ties are present in the data, the likelihood for c becomes flat.

Now, considering that (θ, c) have prior distribution $\pi(\theta, c)$, then the posterior distribution is simply

$$\pi(\theta, c \mid \mathbf{y}, z) \propto f(\mathbf{y} \mid \theta, c)\pi(\theta, c).$$

Note that given the form of the likelihood (16), if θ and c are independent a-priori then they will also be independent a-posteriori. Again, the posterior distribution for λ is the same as before, if we have one period as in (11), or if we have m periods as in (15).

4. Illustration

The reinsurer Swiss Re has been registering information on the most costly insurance losses along the years. The economic research and consulting team produces a bulletin called *sigma* that is published approximately eight times a year, see Swiss Re (2007). The sigma bulletin of natural catastrophes and man-made disasters contains information about the major losses from 1970 up to date. An insurance loss is considered catastrophic if it exceeds \$2,000 millions of US dollars.

Catastrophic losses in the decade of the 1990's and what has been occurred in this decade of 2000's is reported in Table 1. The losses are indexed to 2006 which makes it possible to compare them. We note that in the 1990's there were 18 catastrophic losses whereas up to 2006 there have been 15 with no catastrophic losses that exceeded \$2,000 registered in 2006.

We will carry out a comparison when fitting the traditional CPP and the two ECP models introduced in this paper. The data have been rounded to the nearest hundred to exploit the dependence captured by the $\mathrm{ECP}_{\mathrm{np}}$ with a nonparametric construction when using the Dirichlet process. This rounding has no impact on the other two processes, the CPP and the $\mathrm{ECP}_{\mathrm{p}}$ with parametric construction. Analysis of extreme data usually requires the use of a heavy-tail distribution. Beirlant et al. (1996) suggest the pareto distribution for this kind of data.

Let X_t be the aggregated total catastrophic claims made to an insurance company up to time t such that the number of catastrophic claims N_t follows a homogeneous Poisson process $\mathrm{Po}(\lambda t)$ and the size of the individual claims Y_i are known to exceed \$2,000. The data in Table 1 can be regarded as two independent realizations of the process X_t, such that up to times t_1 and t_2, X_{t_1} and X_{t_2} represent the total catastrophic claims in the decade of 1990 and what has been elapsed in the 2000's, respectively. Let n_1 be the number of claims in the 1990's with sizes $\mathbf{y}_1 = (y_{11}, \ldots, y_{n_1,1})$ and n_2 the number of claims in the 2000's with sizes $\mathbf{y}_2 = (y_{12}, \ldots, y_{n_2,2})$.

We then start by assuming that X_{t_j} is a CPP with independent claims, that is, $Y_{ij} \mid a, b \sim \mathrm{Pa}(a, b)$ are all independent for $i = 1, \ldots, n_j$ and $j = 1, 2$ with pareto distribution and density given by $f(x \mid a, b) = ab^a x^{-(a+1)} I(x \geq b)$. The parameter b in the pareto distribution determines the lower bound for the support of the data. Some authors (see for example, Kaiser and Brazauskas, 2006) fix this lower bound to be the minimum of the observed data, leaving the pareto distribution with only one unknown parameter, a. In this paper, on the other hand, we carry out a full Bayesian analysis for both parameters and let the data determine the best value for b. If our prior knowledge on (a, b) can be represented by $\pi(a) = \mathrm{Ga}(a \mid \alpha_a, \beta_a)$ and $\pi(b) = \mathrm{Ga}(b \mid \alpha_b, \beta_b)$ independent, then the posterior distribution, given the information of the

Decade of 1990's		Decade of 2000's	
Loss	Date	Loss	Date
7200	25/01/1990	2500	06/04/2001
4900	25/02/1990	4100	05/06/2001
8400	27/09/1991	21400	11/09/2001
2500	20/10/1991	2600	06/08/2002
23000	23/08/1992	3500	02/05/2003
2300	11/09/1992	2400	18/09/2003
2700	10/03/1993	8600	11/08/2004
19000	17/01/1994	5500	26/08/2004
3300	17/01/1995	13700	02/09/2004
3300	01/10/1995	3800	06/09/2004
2300	05/09/1996	4000	13/09/2004
2000	04/07/1997	2100	26/12/2004
4400	20/09/1998	2100	19/08/2005
3400	10/09/1999	10400	20/09/2005
4900	22/09/1999	13000	19/10/2005
2300	03/12/1999		
7000	25/12/1999		
2900	27/12/1999		

Table 1: Catastrophic insurance losses in 1990's and 2000's rounded to hundreds (in USD millions). Extracted from Table 8 in Swiss Re, Sigma Bulletin No. 2, 2007, p. 35.

two decades, is characterized by the conditional distributions

$$\pi(a \mid \mathbf{y}, b) \propto a^{\alpha_a+n_1+n_2-1}\left(\frac{b^{n_1+n_2}e^{-\beta_b}}{\prod_{j=1}^{2}\prod_{i=1}^{n_j} y_{ij}}\right)^{a} I(a>0)$$

and

$$\pi(b \mid \mathbf{y}, a) \propto b^{\alpha_b+(n_1+n_2)a-1}e^{-\beta_b b}I(0<b<y_{(1)}),$$

where $y_{(1)}$ denotes the minimum of the y_{ij}'s in both decades.

Now we will assume that X_t is a ECP. To construct an exchangeable sequence with $\mathrm{Pa}(a,b)$ marginal distributions we will consider both, the parametric and the nonparametric approaches. For the parametric method we take a latent Z with conditional distribution $f(z \mid y) = \mathrm{Ipa}(z \mid c, 1/y)$, where $\mathrm{Ipa}(a,b)$ denotes an inverse Pareto distribution with density function given by $f(x \mid a,b) = ab^a x^{a-1} I(0<x<b^{-1})$. It is not difficult to prove that in this case, $f(y \mid z) = \mathrm{Pa}(y \mid a+c, \max(b,z))$ and

$$f(z) = a/(a+c)\mathrm{Ipa}(z \mid c, 1/b) + c/(a+c)\mathrm{Pa}(z \mid a, b). \tag{17}$$

Therefore, the exchangeable sequence with marginals $\mathrm{Pa}(a,b)$, under a parametric approach, is obtained by taking $Y_i \mid Z \sim \mathrm{Pa}(a+c, \max(b,z))$ for $i=1,2,\ldots$ conditionally independent given Z and Z is assigned a mixture distribution given by (17). If $a>2$ then

$$\mathrm{Corr}(Y_i, Y_j) = c(2a+c-2)/(a+c-1)^2, \tag{18}$$

for $i \neq j$.

Remember that we have two independent exchangeable sequences, one for the 1990's and other for the 2000's. For this particular construction, the joint distribution independent of the latent Z, as in (13), has an analytical expression. Therefore, the likelihood, obtained as in (12), using information of the two decades has the form

$$f(y_1,\ldots,y_n \mid a,b,c)=\prod_{j=1}^{2}\left[a(a+c)^{n_j-1}\left(\prod_{i=1}^{n_j}y_{ij}\right)^{-(a+c+1)}\phi_j(a,b,c)I(b\le y_{(1)})\right],$$

where

$$\phi_j(a,b,c)=\left\{b^{n_j(a+c)}+c\,b^a\left(\frac{y_{(1)}^{(n_j-1)a+n_jc}-b^{(n_j-1)a+n_jc}}{(n_j-1)a+n_jc}\right)\right\},$$

for $j=1,2$. Now, assuming that (a,b,c) are independent a-priori such that $\pi(a)=\mathrm{Ga}(a\mid\alpha_a,\beta_a)$, $\pi(b)=\mathrm{Ga}(b\mid\alpha_b,\beta_b)$ and $\pi(c)=\mathrm{Ga}(c\mid\alpha_c,\beta_c)$, then the posterior conditional distributions are,

$$\pi(a\mid\mathbf{y},b,c)\propto a^{\alpha_a+1}(a+c)^{n_1+n_2-2}\left(e^{\beta_a}\prod_{j=1}^{2}\prod_{i=1}^{n_j}y_{ij}\right)^{-a}\left\{\prod_{j=1}^{2}\phi_j(a,b,c)\right\}I(a>0),$$

$$\pi(b\mid\mathbf{y},a,c)\propto b^{\alpha_b-1}e^{-\beta_b b}\left\{\prod_{j=1}^{2}\phi_j(a,b,c)\right\}I(0<b\le y_{(1)})$$

and

$$\pi(c\mid\mathbf{y},a,b)\propto c^{\alpha_c-1}(a+c)^{n_1+n_2-1}\left(e^{\beta_c}\prod_{j=1}^{2}\prod_{i=1}^{n_j}y_{ij}\right)^{-c}\left\{\prod_{j=1}^{2}\phi_j(a,b,c)\right\}I(c>0).$$

For the construction of an exchangeable sequence with $\mathrm{Pa}(a,b)$ marginals under the nonparametric approach we will use the Dirichlet process as in Example 2. That is, we take $Y_i\mid G\sim G$ for $i=1,2,\ldots$ conditionally independent given G and $G\sim\mathcal{DP}(F/c)$ with $F(\mathrm{d}y)=\mathrm{Pa}(y\mid a,b)\mathrm{d}y$. Due to the nonparametric nature of the dependence, $\mathrm{Corr}(Y_i,Y_j)=c/(c+1)$ for $i\ne j$ as in the gamma case of Example 2.

If we use the same prior distribution for (a,b,c) as in the parametric construction, given the information of the two decades, the conditional posterior distributions required to make inference become

$$\pi(a\mid\mathbf{y},b,c)\propto a^{\alpha_a+k_1+k_2-1}\left(\frac{b^{k_1+k_2}e^{-\beta_a}}{\prod_{j=1}^{2}\prod_{i=1}^{k_j}y_{ij}^*}\right)^{a}I(a>0),$$

where, k_1 and k_2 denote the number of distinct observations in each sample and $(y_{11}^*,\ldots,y_{k_1,1}^*)$ and $(y_{12}^*,\ldots,y_{k_2,2}^*)$ are the distinct observations, respectively.

$$\pi(b\mid\mathbf{y},a,c)\propto b^{\alpha_b+(k_1+k_2)a-1}e^{-\beta_b b}I(0\le b\le y_{(1)}),$$

and

$$\pi(c\mid\mathbf{y},a,b)\propto\frac{\Gamma^2(1/c)}{\Gamma(1/c+n_1)\Gamma(1/c+n_2)}\,c^{\alpha_c+k_1+k_2-1}e^{-\beta_c c}I(c>0).$$

Finally, in all cases, the posterior distribution for the intensity parameter λ of the Poisson process $N(t)$, when considering a conjugate analysis, is given by equation (15) with $m=2$.

We carried out posterior inference for the catastrophic insurance losses by implementing a Gibbs sampler for the three models, CPP, $\mathrm{ECP_p}$ and $\mathrm{ECP_{np}}$. To avoid numerical problems we scaled the data by 1000. We considered vague prior distributions for a, b and λ, i.e., we took $(\alpha_a, \beta_a) = (0.01, 0.01)$, $(\alpha_b, \beta_b) = (0.01, 0.01)$ and $(\alpha_\lambda, \beta_\lambda) = (0.01, 0.01)$. Our main parameter of interest is the one that controls the dependence among observations, that is c. We took different values of the hyper-parameters of $\pi(c)$, and in order to compare the fitting of the models to the data, we use the logarithm of the pseudo-marginal likelihood (LPML) statistic. This statistic is a measure of the marginal fitting to the data and has been used as a model selection criterion in many different contexts (see, for example, Sinha and Dey, 1997). Table 2 summarizes the LPML statistic for different values of the hyper-parameters (α_c, β_c) of the prior distribution of c. The Gibbs sampler was run for 50,000 iterations with a burn-in period of 10,000, taking the last 40,000 for computing the estimates.

α_c	β_c	LPML		
		CPP	ECP_p	ECP_np
0.01	0.01	-79.84	-80.05	-80.39
0.1	0.1	-79.82	-79.88	-80.43
1	1	-80.03	-82.52	-80.35
2	0.1	-79.86	-115.60	-80.24

Table 2: Monte Carlo estimates of LPML for different values of (α_c, β_c).

As we can see from Table 2, the fitting of the three models is roughly the same for the first three specifications of the hyper-parameters (α_c, β_c). When we take a slightly informative prior by setting $(\alpha_c, \beta_c) = (2, 0.1)$, the fitting of the ECP_p model becomes worse, while the fitting of the other two is still as good as before. The CPP and the ECP_np seem to be quite robust to the choice of the prior distribution of c.

Parameter	Model	Mean	Median	95% CI
a	CPP	1.180	1.175	(0.803, 1.578)
	ECP_p	1.136	1.142	(0.780, 1.600)
	ECP_np	1.046	1.040	(0.693, 1.456)
b	CPP	1.949	1.964	(1.807, 1.998)
	ECP_p	1.951	1.966	(1.817, 1.998)
	ECP_np	1.932	1.952	(1.746, 1.999)
c	ECP_p	0.035	0.0003	(0.0001, 0.332)
	ECP_np	0.026	0.023	(0.008, 0.059)

Table 3: Posterior summaries of (a, b, c) for the three models.

In order to make a fair comparison among the different models, we considered the case when the fitting of the three models is the same, which corresponds to taking $(\alpha_c, \beta_c) = (0.1, 0.1)$ and vague priors for the rest of the parameters. Posterior summaries for the parameters in the models are presented in Table 3.

¿From Table 3 we can observe that posterior estimates of a and b are consistent in the three models, this is a consequence of the same measure of marginal fitting given in the LPML statistic in Table 2. However, posterior estimates of the parameter c, that controls the dependence, are different. This is to be expected given the difference in the definition of the exchangeable sequence in each model. Figure 3 shows the posterior distribution of c for the two constructions. The posterior distribution of c in the parametric case (left panel) presents a larger dispersion than the posterior distribution of c in the nonparametric construction (right panel).

A better way of appreciating the impact of the parameter c in both exchangeable sequences,

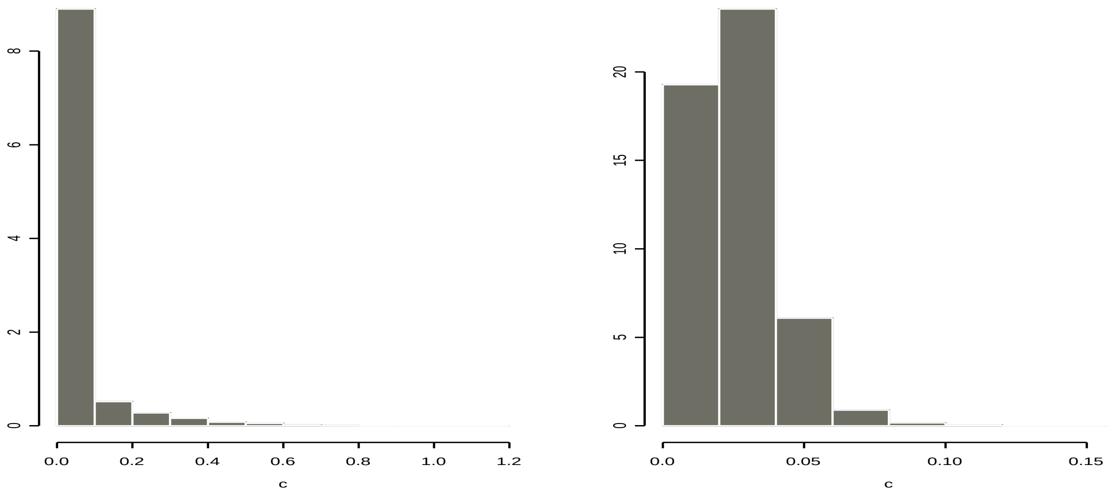

Figure 3: Posterior distribution of c: (Left panel) parametric construction and (right panel) nonparametric construction.

is by computing a measure of dependence. The correlation coefficient (18) for the parametric case is only defined for $a > 2$, and looking at the summaries of the posterior distribution of a in Table 3, we can see that the upper limit of the 95% credible interval is 1.6, therefore the correlation coefficient is not defined for this data set. However, for the nonparametric case, the correlation coefficient given in Proposition 2, can certainly be computed. In addition, we also computed the Kendall's tau measure of dependence based on a bivariate sample from the predictive distribution of both ECP models. These measures are included in Table 4.

Measure	ECP_p	ECP_np
ρ	–	0.025
τ	0.036	0.030

Table 4: Measures of dependence for ECP models: Correlation coefficient ρ and Kendall's tau τ.

If we look at the Kendall's tau measures reported in table 4 we can see that they are similar in both ECP models, 0.036 for the parametric and 0.03 for the nonparametric. These values suggest that the dependence among these catastrophic claims within each decade is weak. Despite the small degree of dependence, the introduction of the extra parameter c in the ECP models does have an impact in the predictive distributions of a single claim and in the whole process X_t. A log-normal smoothing of the marginal predictive distributions of a future claim, say Y_F, for the three models are shown in Figure 4. Although in the graph the three lines are almost indistinguishable, numerically we can see that the predictive distributions when assuming exchangeability have heavier tails. The 95% upper quantiles (in the original units) are 22759, 24228 and 27490 for the CPP, ECP_p and ECP_np models respectively.

Finally, we carried out a predictive analysis for the aggregated claims for the rest of the decade of 2000's, that is from 2007 to the end of 2009. For that we needed to consider the information about the frequencies of occurrence of the claims, modeled by the Poisson process N_t and in particular by the intensity λ. Considering that in 1990's we observed 18 claims with the last claim made on 27/12/2009, and in the 2000's we have observed 15 claims with the last one made on 19/10/2005. Then the posterior distribution for λ, under the vague prior conditions

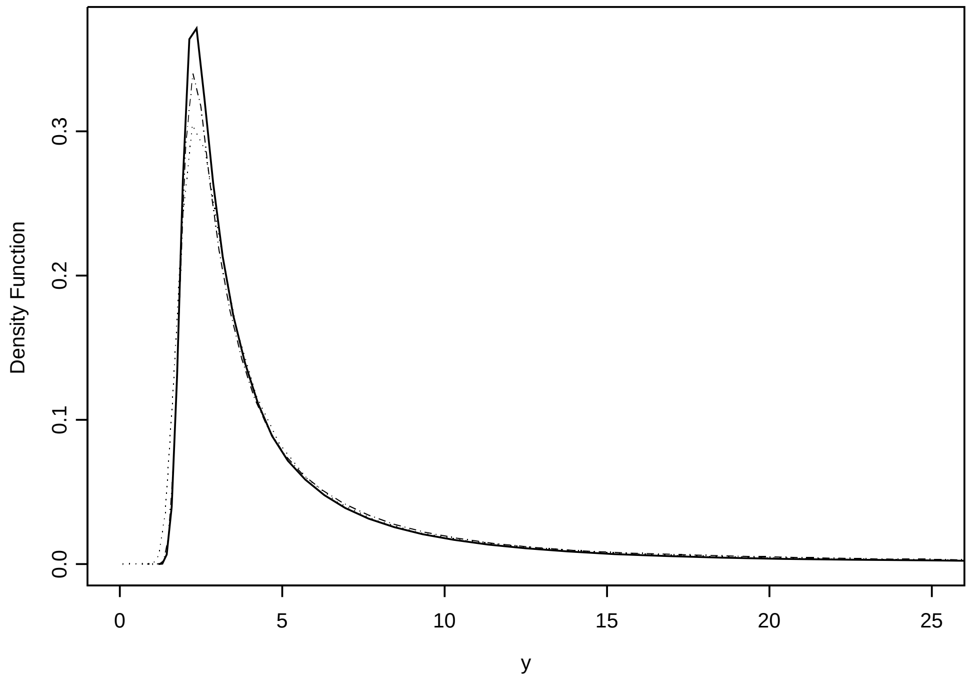

Figure 4: Posterior predictive distribution for a single claim Y_F in thousands. (—) CPP, ($\cdots$) $\mathrm{ECP_p}$ and (-·-) $\mathrm{ECP_{np}}$.

mentioned above, is Ga(33.01, 15.79). Therefore, the posterior mean rate is 2.09 claims per year.

With the posterior distribution for λ and the posterior predictive distribution for the whole sequence of claims, we obtained the predictive distribution for the future catastrophic claims in the last three years of the decade, say X_3. A log-normal smoothing of this predictive distribution is included in Figure 5 for the three cases, CPP (solid line), $\mathrm{ECP_p}$ (dotted line) and $\mathrm{ECP_{np}}$ (dashed-dotted line). In this Figure we can observe that the predictive distribution for the future claims is overdispersed for both ECP models with respect to the CPP model. For the insurance company, having slightly heavier tails represents a more conservative scenario because the model is giving more probability to large claims, and this is something important when dealing with catastrophic data. The insurance companies covering these catastrophic events must reserve certain amount of money large enough to cover possible losses. A conservative approach could be to take the 95% quantile of the distribution of future claims to determine the reserve. These quantiles are plotted as vertical lines in Figure 5 whose values are (in the original units) 200604, 215413 and 290066 for the CPP, $\mathrm{ECP_p}$ and $\mathrm{ECP_{np}}$ respectively. Certainly the reserves obtained from any of the ECP models give more warranties to the insurers.

5. Discussion.

In this paper we introduced a generalization of the compound Poisson process by relaxing the independence assumption in the claims to a more realistic exchangeable (positive correlated) assumption. By doing that, the resulting process turns out to ruin faster than when using the traditional CPP process. Moreover, predictive distributions of the aggregated claims are more conservative, assigning a bigger probability to large claims.

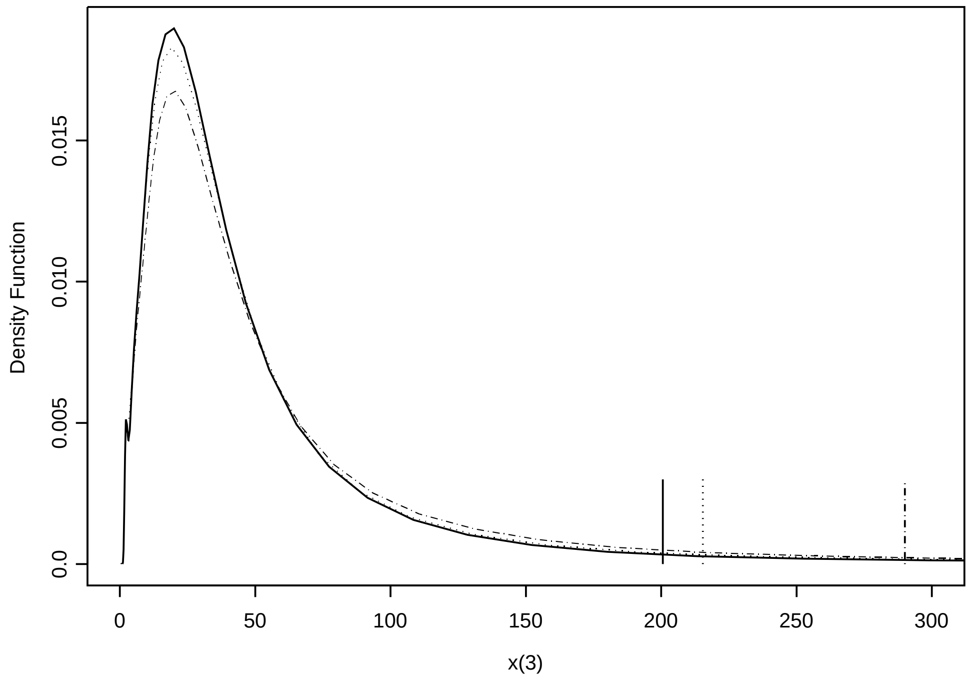

Figure 5: Posterior predictive distribution for the total amount of claims in the last three years of the decade of 2000's, X_3, in thousands. (—) CPP, ($\cdots$) $\mathrm{ECP}_{\mathrm{p}}$ and (-·-) $\mathrm{ECP}_{\mathrm{np}}$. The vertical lines correspond to the 95% quantiles, respectively.

We proposed two novel and general ways of defining exchangeable sequences with any particular marginal distribution, by means of conditioning on a latent random variable (parametric approach) or a latent random distribution (nonparametric approach). We illustrated our proposed methodology to construct exchangeable sequences for the gamma and pareto cases, but it can certainly be applied to any distribution.

We dealt in detail with the problem of estimating the parameters of the compound Poisson process and the exchangeable claims process and for that we followed a Bayesian estimation approach. Estimability issues were also considered and addressed appropriately by considering several samples.

References

Beirlant, J., Teugels, J.L. and Vynckier, P., 2006. *Practical analysis of extreme values.* Leuven: Leuven University Press.

Blackwell, D. and MacQueen, J.B., 1973. Ferguson distributions via pólya urn schemes. *Annals of Statistics*, **1**, 353-355.

Bühlmann, H., 1960. *Austauschbare stochatische Variabeln unf ihre Grenzwertsaetze.* Univ. of California Press.

Bühlmann, H., 1970. *Mathematical Methods in Risk Theory.* Springer.

Cossette, H. and Marceau, E., 2000. The discrete-time risk model with correlated classes of business. *Insurance: Mathematics and Economics*, **26**, 133-149.

Cramér, H., 1930. *On the mathematical theory of risk.* Skandia Jubilee Volume. Stockholm.

Daboni, D., 1974. Some models of inference in the risk theory from a Bayesian viewpoint. *ASTIN bull.*, **8**, 38-56.

de Finetti, B., 1937. Le prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré (Paris)*, **7**, 1–68.

Ferguson, T., 1973. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209-230.

Feller, W., 1971. *An Introduction to Probability Theory and Its Applications, Volume 2.* Wiley.

Gerber, H. U., 1982. Ruin theory in the linear model. *Insurance: Mathematics and Economics*, **1**, 177-184.

Grandell, J., 1997. *Mixed Poisson Processes.* Chapman and Hall.

Kaiser, T. and Brazauskas, V., 2006. Interval estimation of actuarial risk measures. *North American Actuarial Journal*, **10**, 249–268.

Klass, M. and Teicher, H., 1987. The central limit theorem for exchangeable random variables without moments. *The Annals of Probability*, **15**, 138–153.

Klugman, S. A., 2003. *Bayesian Statistics in Actuarial Science: With emphasis on credibility.* New York: springer.

Lundberg, F., 1926. *Försäkringsteknisk Riskutjämning.* F. Englunds boktryckeri A.B., Stockholm.

Mena, R. H. and Walker, S. G., 2005. Stationary autoregressive models via a Bayesian nonparametric approach. *Journal of Time Series Analysis*, **26**, 789-805.

Mikosch,T. and Samorodnitsky, G., 2000. Ruin probability with claims modeled by a stationary ergodic stable process. *The Annals of Probability*, **28**, 1814-1854.

Morales, M. and Schoutens, W., 2003. A risk model driven by Lévy processes. *Applied Stochastic Models in Business and Industry*, **19**, 147–167.

Pitman, J., 1996. Some developments of the Blackwell-MacQueen urn scheme. *Probability and Game Theory. Papers in honor of David Blackwell*, (ed T.S. Ferguson *et al.*), IMS Lecture Notes Monograph Series, **30**, 245–267.

Pitt, M. K. and Walker, S. G., 2005. Constructing stationary time series models using auxiliary variables with applications. *Journal of the American Statistical Association*, **100**, 554–564.

Pitt, M. K., Chatfield, C., and Walker, S. G., 2002. Constructing first order autoregressive models via latent processes. *Scandinavian Journal of Statistics*, **29**, 657–663.

Promislow, S. D., 1991. The probability of ruin in a process with dependent increments. *Insurance: Mathematics and Economics*, **10**, 99-107.

Regazzini, E., Lijoi, A. and Prünster, I., 2003. Distributional results for means of random measures with independent increments. *Annals of Statistics*, **31**, 560–585.

Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J., 2001. *Stochastic processes for insurance and finance*. New York: Wiley.

Sato, K. I., 1999. *Lévy processes and infinitely divisible distributions.* Cambridge University Press.

Sinha, D. and Dey, D. K., 1997. Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, **92**, 1195–1212.

Smith, A. F. M. and Roberts, G. O., 1993. Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* B, **55**, 3–23.

Swiss Re, 2007. Natural catastrophes and man-made disasters 2006: low insured losses. *Sigma*, **2**/2007. Available at www.swissre.com.

Taylor, H. M. and Karlin, S., 1998. *An Introduction to Stochastic Modeling.* Academic Press.